

VOICE RECOGNIZING METHOD AND DEVICE

Publication number: JP2001005482

Publication date: 2001-01-12

Inventor: KUNIEDA NOBUYUKI

Applicant: MATSUSHITA ELECTRIC IND CO LTD

Classification:

- international: **G10L15/10; G10L15/06; G10L15/18; G10L15/28; G10L15/00;** (IPC1-7): G10L15/10; G10L15/18; G10L15/28

- European: G10L15/06A3

Application number: JP19990173818 19990621

Priority number(s): JP19990173818 19990621

Also published as:



EP1063634 (A2)

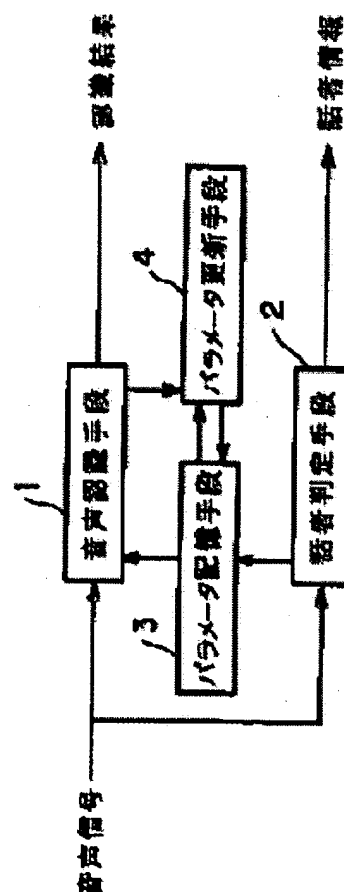
EP1063634 (A3)

Report a data error here

Abstract of JP2001005482

PROBLEM TO BE SOLVED: To output voice recognition results of high performance when voices by plural speakers are inputted.

SOLUTION: This recognizing method is provided with a speaker judging means 2 judging whether a speaker who uttered a voice is a speaker who is recognized at the past or a speaker who is recognized for the first time from an inputted voice and also transfers results processed in the means 2 to a parameter storage means 3 to produce recognition parameters for every speaker and enables the recognition parameters so as to be stored. Moreover, the recognition parameters for every speaker stored in the storage means 3 are made adaptive in a parameter updating means 4 so as to become optimum recognition parameters to respective speakers by repeating voice recognizing processings in a voice recognizing means 1.



Data supplied from the esp@cenet database - Worldwide

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号
特開2001-5482
(P2001-5482A)

(43) 公開日 平成13年1月12日 (2001.1.12)

(51) Int.Cl.⁷

識別記号

F I

テ-マ-ト* (参考)

G 1 0 L 15/10
15/28
15/18

G 1 0 L 3/00

5 3 1 K 5 D 0 1 5
5 1 1
5 3 7 Z

審査請求 有 請求項の数15 O L (全 9 頁)

(21) 出願番号 特願平11-173818

(22) 出願日 平成11年6月21日 (1999.6.21)

(71) 出願人 000003821

松下電器産業株式会社

大阪府門真市大字門真1006番地

(72) 発明者 國枝 伸行

神奈川県横浜市港北区綱島東四丁目3番1
号 松下通信工業株式会社内

(74) 代理人 100082692

弁理士 蔵合 正博

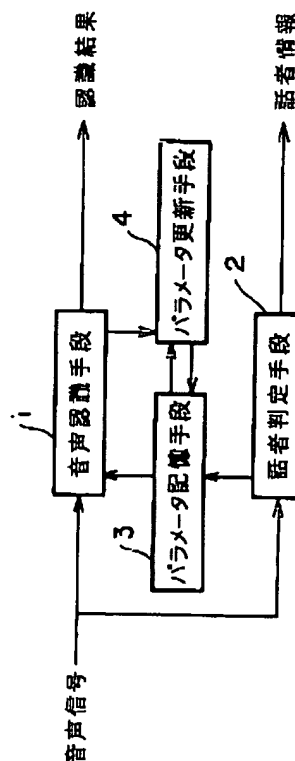
Fターム (参考) 5D015 AA03 DD02 HH00 HH03 HH07
HH21 KK04

(54) 【発明の名称】 音声認識方法及び装置

(57) 【要約】

【課題】 複数の話者が発声した音声が入力された場合に、より高性能な音声認識結果を出力する。

【解決手段】 入力された音声から音声が発話した話者が過去に認識した発話者であるのか、初めて認識をする発話者であるかを判定する話者判定手段2を備えるとともに、この話者判定手段2で処理された結果をパラメータ記憶手段3に渡し、話者ごとに認識パラメータを作成して記憶できるようにする。また、パラメータ記憶手段3に記憶された各話者ごとの認識パラメータは、音声認識手段1で音声認識処理を繰り返すことによって各話者に最適な認識パラメータとなるようにパラメータ更新手段4で適応化される。



【特許請求の範囲】

【請求項1】 音声信号から発声した内容及び発話者を認識する音声認識方法であって、入力された音声信号から発話者が過去に認識した話者であるか新規に認識する話者であるかを判定するステップと、前記発話者が新規に認識する話者であると判定された場合に、その発話者と認識パラメータとを関連づけて記憶し、出力するとともに、前記発話者が過去に認識した話者であると判定された場合に、その話者と関連づけられた認識パラメータを出力するステップと、前記入力された音声信号と前記出力された認識パラメータとから音声認識処理を行い、認識結果を出力するステップと、前記音声認識結果から前記認識パラメータを逐次更新するステップとを備えたことを特徴とする音声認識方法。

【請求項2】 前記音声認識結果と前記話者判定結果とから言語的なつながりを考慮して認識結果を補正して出力するステップを備えたことを特徴とする請求項1記載の音声認識方法。

【請求項3】 前記音声認識結果と前記話者判定結果とから合成音を生成して出力するステップを備えたことを特徴とする請求項1記載の音声認識方法。

【請求項4】 複数のマイクロホンから入力された音声信号から発声した内容及び発話者を認識する音声認識方法であって、前記複数のマイクロホンから入力された信号から目的の音声信号を得るステップと、前記得られた音声信号の到来方向を判定するステップと、前記得られた音声信号と前記信号方向判定結果から発話者が過去に認識した話者であるか新規に認識する話者であるかを判定するステップと、前記発話者が新規に認識する話者であると判定された場合に、その発話者と認識パラメータとを関連づけて記憶し、出力するとともに、前記発話者が過去に認識した話者であると判定された場合に、その話者と関連づけられた認識パラメータを出力するステップと、前記得られた音声信号と前記出力された認識パラメータとから音声認識処理を行い、認識結果を出力するステップと、前記音声認識結果から前記認識パラメータを逐次更新するステップとを備えたことを特徴とする音声認識方法。

【請求項5】 前記音声認識結果と前記話者判定結果とから言語的なつながりを考慮して認識結果を補正して出力するステップを備えたことを特徴とする請求項4記載の音声認識方法。

【請求項6】 前記音声認識結果と前記話者判定結果とから合成音を作成して出力するステップを備えたことを特徴とする請求項4記載の音声認識方法。

【請求項7】 音声信号から発声した内容及び発話者を認識する音声認識装置であって、入力された音声信号から発話者が過去に認識した話者であるか新規に認識する話者であるかを判定する話者判断手段と、前記発話者が新規に認識する話者であると判定された場合に、その発

話者と認識パラメータとを関連づけて記憶し、出力するとともに、前記発話者が過去に認識した話者であると判定された場合に、その話者と関連づけられた認識パラメータを出力するパラメータ記憶手段と、前記入力された音声信号と前記パラメータ記憶手段から出力された認識パラメータとから音声認識処理を行い、認識結果を出力する音声認識手段と、前記音声認識手段で認識処理を行った結果から前記パラメータ記憶手段に記憶されている認識パラメータを逐次更新するパラメータ更新手段とを備えたことを特徴とする音声認識装置。

【請求項8】 前記音声認識手段で処理した音声認識結果と前記話者判定手段で抽出された発話者に関する情報から言語的なつながりを考慮して認識結果を補正して出力する認識結果補正手段を備えたことを特徴とする請求項7記載の音声認識装置。

【請求項9】 前記音声認識手段で処理した音声認識結果と前記話者判定手段で抽出された話者判定結果とから合成音を生成して出力する音声合成手段、を備えたことを特徴とする請求項7記載の音声認識装置。

【請求項10】 複数のマイクロホンから入力された音声信号から発声した内容及び発話者を認識する音声認識装置であって、前記複数のマイクロホンから入力された信号から目的の音声信号を得るビームフォーマと、前記ビームフォーマから音声信号の到来方向を判定する信号方向判定手段と、前記ビームフォーマで得られた音声信号と前記信号方向判定手段で得られた結果とから発話者が過去に認識した話者であるか新規に認識する話者であるかを判定する話者判定手段と、前記発話者が新規に認識する話者であると判定された場合に、その発話者と認識パラメータとを関連づけて記憶し、出力するとともに、前記発話者が過去に認識した話者であると判定された場合に、その話者と関連づけられた認識パラメータを出力するパラメータ記憶手段と、前記ビームフォーマから得られた音声信号と前記パラメータ記憶手段から出力された認識パラメータとから音声認識処理を行い、認識結果を出力する音声認識手段と、前記音声認識手段で認識処理を行った結果から前記パラメータ記憶手段に記憶されている認識パラメータを逐次更新するパラメータ更新手段とを備えたことを特徴とする音声認識装置。

【請求項11】 前記音声認識手段で処理した音声認識結果と前記話者判定手段で抽出された話者判定結果とから言語的なつながりを考慮して認識結果を補正して出力する認識結果補正手段を備えたことを特徴とする請求項10記載の音声認識装置。

【請求項12】 前記音声認識手段で処理した音声認識結果と前記話者判定手段で抽出された話者判定結果とから合成音を作成して出力する音声合成手段を備えたことを特徴とする請求項10記載の音声認識装置。

【請求項13】 請求項7から12のいずれかに記載の音声認識装置において、話者情報と関連づけた認識パラ

メータとして標準パターンを選択することを特徴とする音声認識装置。

【請求項14】 請求項7から12のいずれかに記載の音声認識装置において、音声認識処理手段にDPマッチング処理を利用することを特徴とする音声認識装置。

【請求項15】 請求項7から12のいずれかに記載の音声認識装置において、音声認識処理手段に隠れマルコフモデルを利用することを特徴とする音声認識装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、音声認識の方法及び装置に関するものであり、特に複数の話者が発声した音声进行認識することに有効な音声認識方法及び装置に関する。

【0002】

【従来の技術】通常の音声認識方法及び装置では、入力された音声に対して認識処理を行い、認識結果を決定して出力する。この場合、発話者の種類、感情、発話速度などの発話者に関する情報は取り除いて処理され、複数の話者が発声した場合でもすべて同じ認識パラメータを利用して音声認識が行われる。

【0003】一方、入力された音声の質の変化に応じて認識パラメータを選択することも試みられている。例えば、特開平7-56595号公報記載の発明では、発話者の様態の変化や話者の変化を検出する手段を有し、発話音声を分析した結果を利用して照合手段あるいは判定手段を制御するように構成され、発話された音声の変化に対応して音声認識する。

【0004】

【発明が解決しようとする課題】しかしながら、上記従来の方法では、複数の話者が交互に発声している場合、男声用認識パラメータを利用するか女声用認識パラメータを利用するかという選択をすることは可能であっても、複数の話者それぞれに専用の認識パラメータを準備して認識処理を行えるようには構成されていないという問題があった。

【0005】本発明は、このような問題を解決するためになされたもので、複数の話者が発声した音声が入力された場合に、より高性能な音声認識結果を出力する音声認識方法及び装置を提供するものである。

【0006】

【課題を解決するための手段】本発明の音声認識方法は、音声信号から発声した内容及び発話者を認識する音声認識方法であって、入力された音声信号から発話者が過去に認識した話者であるか新規に認識する話者であるかを判定するステップと、前記発話者が新規に認識する話者であると判定された場合に、その発話者と認識パラメータとを関連づけて記憶し、出力するとともに、前記発話者が過去に認識した話者であると判定された場合に、その話者と関連づけられた認識パラメータを出力す

るステップと、前記入力された音声信号と前記出力された認識パラメータとから音声認識処理を行い、認識結果を出力するステップと、前記音声認識結果から前記認識パラメータを逐次更新するステップとを備えたことを特徴とした構成を有している。この構成により、高性能な音声認識が可能となる。

【0007】また、本発明の音声認識方法は、前記音声認識結果と前記話者判定結果とから言語的なつながりを考慮して認識結果を補正して出力するステップを備えたことを特徴とした構成を有している。この構成により、高性能な音声認識が可能となると同時に、発話者に関する情報が同一である直前に発声された内容を考慮して認識結果を補正することが可能となる。

【0008】また、本発明の音声認識方法は、前記音声認識結果と前記話者判定結果とから合成音を生成して出力するステップを備えたことを特徴とした構成を有している。この構成により、高性能な音声認識が可能となると同時に、認識結果に話者情報を付加して合成音として出力することが可能となる。

【0009】また、本発明の音声認識方法は、複数のマイクロホンから入力された音声信号から発声した内容及び発話者を認識する音声認識方法であって、前記複数のマイクロホンから入力された信号から目的の音声信号を得るステップと、前記得られた音声信号の到来方向を判定するステップと、前記得られた音声信号と前記信号方向判定結果から発話者が過去に認識した話者であるか新規に認識する話者であるかを判定するステップと、前記発話者が新規に認識する話者であると判定された場合に、その発話者と認識パラメータとを関連づけて記憶し、出力するとともに、前記発話者が過去に認識した話者であると判定された場合に、その話者と関連づけられた認識パラメータを出力するステップと、前記得られた音声信号と前記出力された認識パラメータとから音声認識処理を行い、認識結果を出力するステップと、前記音声認識結果から前記認識パラメータを逐次更新するステップとを備えたことを特徴とした構成を有している。この構成により、複数の話者が発声した音声が入力された場合でも、高性能な音声認識が可能となる。

【0010】また、本発明の音声認識方法は、前記音声認識結果と前記話者判定結果とから言語的なつながりを考慮して認識結果を補正して出力するステップを備えたことを特徴とした構成を有している。この構成により、高性能な音声認識が可能となると同時に、発話者に関する情報が同一である直前に発声された内容を考慮して認識結果を補正することが可能となる。

【0011】また、本発明の音声認識方法は、前記音声認識結果と前記話者判定結果とから合成音を作成して出力するステップを備えたことを特徴とした構成を有している。この構成により、高性能な音声認識が可能となると同時に、認識結果に話者情報を付加して合成音として

出力することが可能となる。

【0012】また、本発明の音声認識装置は、音声信号から発生した内容及び発話者を認識する音声認識装置であって、入力された音声信号から発話者が過去に認識した話者であるか新規に認識する話者であるかを判定する話者判断手段と、前記発話者が新規に認識する話者であると判定された場合に、その発話者と認識パラメータとを関連づけて記憶し、出力するとともに、前記発話者が過去に認識した話者であると判定された場合に、その話者と関連づけられた認識パラメータを出力するパラメータ記憶手段と、前記入力された音声信号と前記パラメータ記憶手段から出力された認識パラメータとから音声認識処理を行い、認識結果を出力する音声認識手段と、前記音声認識手段で認識処理を行った結果から前記パラメータ記憶手段に記憶されている認識パラメータを逐次更新するパラメータ更新手段とを備えたことを特徴とした構成を有している。この構成により、発話者情報を認識処理に利用して高性能な音声認識処理を行うことが可能となる。

【0013】また、本発明の音声認識装置は、前記音声認識手段で処理した音声認識結果と前記話者判定手段で抽出された発話者に関する情報から言語的なつながりを考慮して認識結果を補正して出力する認識結果補正手段を備えたことを特徴とした構成を有している。この構成により、発話者情報を認識処理に利用して高性能な音声認識処理を行うことが可能となると同時に、発話者に関する情報が同一である直前に発声された内容を考慮して認識結果を補正することが可能となる。

【0014】また、本発明の音声認識装置は、前記音声認識手段で処理した音声認識結果と前記話者判定手段で抽出された話者判定結果とから合成音を生成して出力する音声合成手段を備えたことを特徴とした構成を有している。この構成により、発話者情報を認識処理に利用して高性能な音声認識処理を行うことが可能となると同時に、認識結果に話者情報を付加して合成音として出力することが可能となる。

【0015】また、本発明の音声認識装置は、複数のマイクロホンから入力された音声信号から発声した内容及び発話者を認識する音声認識装置であって、前記複数のマイクロホンから入力された信号から目的の音声信号を得るビームフォーマと、前記ビームフォーマから音声信号の到来方向を判定する信号方向判定手段で得られた手段と、前記ビームフォーマで得られた音声信号と前記信号方向判定結果とから発話者が過去に認識した話者であるか新規に認識する話者であるかを判定する話者判定手段と、前記発話者が新規に認識する話者であると判定された場合に、その発話者と認識パラメータとを関連づけて記憶し、出力するとともに、前記発話者が過去に認識した話者であると判定された場合に、その話者と関連づけられた認識パラメータを出力するパラメータ記憶手

段と、前記ビームフォーマから得られた音声信号と前記パラメータ記憶手段から出力された認識パラメータとから音声認識処理を行い、認識結果を出力する音声認識手段と、前記音声認識手段で認識処理を行った結果から前記パラメータ記憶手段に記憶されている認識パラメータを逐次更新するパラメータ更新手段とを備えたことを特徴とした構成を有している。この構成により、発話者情報を認識処理に利用して高性能な音声認識処理を行うことが可能となる。

【0016】また、本発明の音声認識装置は、前記音声認識手段で処理した音声認識結果と戦記話者判定手段で抽出された話者判定結果とから言語的なつながりを考慮して認識結果を補正して出力する認識結果補正手段を備えたことを特徴とした構成を有している。この構成により、発話者情報を認識処理に利用して高性能な音声認識処理を行うことが可能となると同時に、発話者に関する情報が同一である直前に発声された内容を考慮して、認識結果を補正することが可能となる。

【0017】また、本発明の音声認識装置は、前記音声認識手段で処理した音声認識結果と前記話者判定手段で抽出された話者判定結果とから合成音を作成して出力する音声合成手段を備えたことを特徴とした構成を有している。この構成により、発話者情報を認識処理に利用して高性能な音声認識処理を行うことが可能となると同時に、認識結果に話者情報を付加して合成音として出力することが可能となる。

【0018】また、本発明の音声認識装置は、請求項7から12のいずれかに記載の音声認識装置において、話者情報と関連づけた認識パラメータとして標準パターンを選択することを特徴とした構成を有している。この構成により、発話者情報を認識処理に利用して高性能な音声認識処理を行うことが可能となる。

【0019】また、本発明の音声認識装置は、請求項7から12のいずれかに記載の音声認識装置において、音声認識処理手段にDPマッチング処理を利用することを特徴とした構成を有している。この構成により、発話者情報を認識処理に利用して高性能な音声認識処理を行うことが可能となる。

【0020】また、本発明の音声認識装置は、請求項7から12のいずれかに記載の音声認識装置において、音声認識処理手段に隠れマルコフモデルを利用することを特徴とした構成を有している。この構成により、発話者情報を認識処理に利用して高性能な音声認識処理を行うことが可能となる。

【0021】

【発明の実施の形態】以下、本発明の実施の形態について、図1から図9を用いて説明する。

【0022】(実施の形態1)図1に示すように、本発明の第1の実施の形態の音声認識装置は、入力された音声信号とパラメータ記憶手段3から出力された認識パラ

メータから音声認識処理を行い、認識結果を出力する音声認識手段1と、入力された音声信号から発話者が過去に認識した話者であるか新規に認識する話者であるかを判定する話者判定手段2と、話者判定手段2で得られた結果から発話者と認識パラメータを関連づけて記憶し、音声認識に利用する認識パラメータを出力するパラメータ記憶手段3と、音声認識手段1で認識処理を行った結果を基にパラメータ記憶手段3に記憶されている認識パラメータを逐次更新するパラメータ更新手段4とを備えている。

【0023】次に、上記のように構成された音声認識装置の動作を、図7に示すパラメータ記憶手段3に記憶された内容と、図8に示す処理の流れを参照しながら説明する。まず、音声信号が入力され(ステップS1)、入力された音声信号は、音声認識手段1と話者判定手段2に渡される。次に、話者判定手段2では、音声信号から発話者が「過去に認識した話者の誰か」であるか「新規に認識する話者」であるかを判定する(ステップS2)。話者判定手段2で話者情報を判定した結果、例えば「入力された音声は過去に認識している話者Aのものである」と判定された場合には、話者情報である「話者A」と関連づけられた「認識パラメータA」をパラメータ記憶手段3から抽出し(ステップS3)、音声認識手段1に渡す(ステップS6)。一方、話者判定の結果、「新規に認識する音声(過去に認識された話者ではない)」と判定された場合には、新規に話者情報「発話者X」を付加し、また、パラメータ記憶手段3に新たな「話者Xの認識パラメータ」を作成し(ステップS4)、話者情報「発話者X」と関連づけて記憶するとともに(ステップS5)、音声認識手段1に渡す(ステップS6)。音声認識手段1では、入力された音声信号とパラメータ記憶手段3から渡された認識パラメータを利用して音声認識処理を行い、認識結果を出力する(ステップS7)。最後に、音声認識手段1で認識処理された結果は、パラメータ更新手段4に渡され、パラメータ記憶手段3に記憶されている認識パラメータを、同一の話者情報を有する音声の認識性能を高めるように更新する(ステップS8)。

【0024】このように、本実施の形態1によれば、入力された音声から発話した話者が過去に認識した発話者であるか、初めて認識をする発話者であるかを判定する話者判定手段2を備えるとともに、この話者判定手段2で処理された結果をパラメータ記憶手段3に渡し、話者ごとに認識パラメータを作成して記憶できるように構成し、さらに、パラメータ記憶手段3に記憶された各話者ごとの認識パラメータを、音声認識手段1で音声認識処理を繰り返すことによって各話者に最適な認識パラメータとなるようにパラメータ更新手段4で適応化するようにしたので、話者ごとに最適な認識パラメータを準備することが可能となり、より高い認識性能を持つ音声認識

を実現できる。

【0025】(実施の形態2)図2は本発明の第2の実施の形態の音声認識装置を示し、図1に示した実施の形態1の構成と異なるのは、認識結果補正手段5を追加したことだけであり、同一の構成要素には同一の符号を付して、重複した説明は省略する。

【0026】入力された音声信号は、音声認識手段1と話者判定手段2に渡される。話者判定手段2では、実施の形態1と同様に音声信号から発話者の判定が行われる。話者判定手段2で判定された結果は、パラメータ記憶手段3と認識結果補正手段5に渡される。パラメータ記憶手段3では、話者情報と関連づけられた認識パラメータを選択して音声認識手段1に渡す。音声認識手段1では、入力された音声信号とパラメータ記憶手段3で選択された認識パラメータを利用して音声認識処理を行い、認識結果補正手段5に渡される。認識結果補正手段5では、音声認識結果と話者情報から同一の話者情報を持つ音声の前後のつながりを分析し、必要に応じて音声認識結果を補正して認識結果を出力する。音声認識手段1で認識処理された結果は、パラメータ更新手段4に渡され、認識パラメータの更新を行う。

【0027】話者に応じて認識結果を補正する方法の例を図9に用いて説明する。図9は4単語を連続して単語音声認識した時の話者情報と認識候補の例を表に示してまとめたものである。この表を見ると、認識順で3番目の単語は話者Aが発声し、認識第1候補は「ボウリング場」となっている。ところが、話者Aは認識順1番目の音声の発声を行っており、その時の認識結果は「ラジオをつけて」という結果であった。この結果を考慮して、再度3番目の発声単語の認識候補を見てみると、第2候補である「ボリューム小さく」である可能性が高くなることが分かる。このような場合、認識結果補正手段5では、同じ話者が前に発声した認識結果を参考にして、認識候補を修正する処理を行う。

【0028】このように、本実施の形態2によれば、実施の形態1の構成に加えて認識結果補正手段5を備えているので、高性能な音声認識が可能となると同時に、発話者に関する情報が同一である直前に発声された内容を考慮して認識結果を補正することが可能となる。

【0029】(実施の形態3)図3は本発明の第3の実施の形態の音声認識装置を示し、図1に示した実施の形態1の構成と異なるのは、音声合成手段6を追加したことだけであり、同一の構成要素には同一の符号を付して重複した説明は省略する。

【0030】入力された音声信号は、音声認識手段1と話者判定手段2に渡される。話者判定手段2では、実施の形態1と同様に音声信号から発話者の判定が行われる。話者判定手段2で判定された発話者情報は、パラメータ記憶手段3と音声合成手段6に渡される。パラメータ記憶手段3では、話者情報にあわせた認識パラメータ

を選択して音声認識手段1に渡す。音声認識手段1では、入力された音声信号とパラメータ記憶手段3から渡された認識パラメータを利用して音声認識処理を行い、音声合成手段6に渡される。音声合成手段6では、音声認識結果と話者情報を統合して認識結果を合成音として出力する。そして最後に、音声認識手段1で認識処理された結果は、パラメータ更新手段4に渡され、話者情報と認識パラメータの対応づけを更新し、パラメータ記憶手段3に渡される。

【0031】このように、本実施の形態3によれば、実施の形態1の構成に加えて、音声合成手段6を備えているので、複数の話者が発声した音声が入力された場合でも、高性能な音声認識が可能となると同時に、認識結果に話者情報を付加して合成音として出力することが可能となる。

【0032】(実施の形態4)図4は本発明の第4の実施の形態の音声認識装置を示し、図1に示した実施の形態1の構成と異なるのは、複数の発話者の音声を入力するマイクロホンアレイ7と、入力された音声信号から目的の音声信号を得るビームフォーマ8と、ビームフォーマ8からの音声信号の到来方向を判定する信号方向判定手段9とを追加したことだけであり、同一の構成要素には同一の符号を付して重複した説明は省略する。

【0033】複数の話者からの音声信号は、複数のマイクロホンから構成されるマイクロホンアレイ7を通じて入力され、ビームフォーマ8に入る。ビームフォーマ8では、目的の音声信号を抽出し、音声認識手段1、話者判定手段2及び信号方向判定手段9に渡す。信号方向判定手段9では、発話者の方向を推定し、その情報を話者判定手段2に渡す。話者判定手段2では、音声信号及び推定した話者方向から、実施の形態1と同様にして、複数の発話者の中から特定の発話者の判定が行われ、話者情報として出力する。以下の処理は実施の形態1とほぼ同じ処理が行われる。すなわち、話者判定手段2で分析された発話者情報は、パラメータ記憶手段3に渡される。さらにパラメータ記憶手段3では、話者情報に合わせた認識パラメータを選択して音声認識手段1に渡す。音声認識手段1では、入力された音声信号とパラメータ記憶手段3から渡された認識パラメータを利用して音声認識処理を行い、認識結果として出力する。最後に、音声認識手段1で認識処理された結果は、パラメータ更新手段4に渡され、話者情報と認識パラメータの対応づけを更新し、パラメータ記憶手段3に渡される。

【0034】このように、本実施の形態4によれば、実施の形態1の構成に加えて、マイクロホンアレイ7、ビームフォーマ8、信号方向判定手段9を備えているので、複数の話者が発声した音声が入力された場合でも、高性能な音声認識が可能となる。

【0035】(実施の形態5)図5は本発明の第5の実施の形態の音声認識装置を示し、図4に示した実施の形

態4の構成と異なるのは、認識結果補正手段5を追加したことだけであり、同一の構成要素には同一の符号を付して、重複した説明は省略する。

【0036】複数の話者からの音声信号は、複数のマイクロホンから構成されるマイクロホンアレイ7を通じて入力され、ビームフォーマ8に入る。ビームフォーマ8では、目的の音声信号を抽出して音声認識手段1、話者判定手段2及び信号方向判定手段9に渡される。信号方向判定手段9では、話者の方向を推定し、その情報を話者判定手段2に渡す。話者判定手段2では、音声信号及び推定した話者方向から特定の発話者の判定が行われる。以下の流れは、実施の形態2と同じである。話者判定手段2で分析された発話者情報は、パラメータ記憶手段3と認識結果補正手段5に渡される。パラメータ記憶手段3では、話者情報に合わせた認識パラメータを選択して音声認識手段1に渡す。音声認識手段1では、入力された音声信号とパラメータ記憶手段3から渡された認識パラメータを利用して音声認識処理を行い、その結果を認識結果補正手段5に渡す。認識結果補正手段5では、音声認識結果と話者情報を統合して認識結果として出力する。最後に、音声認識手段1で認識処理された結果は、パラメータ更新手段4に渡され、話者情報と認識パラメータの対応づけを更新し、パラメータ記憶手段3に渡される。

【0037】このように、本実施の形態5によれば、実施の形態4の構成に加えて、認識結果補正手段5を備えているので、高性能な音声認識が可能となると同時に、発話者に関する情報が同一である直前に発声された内容を考慮して認識結果を補正することが可能となる。

【0038】(実施の形態6)図6は本発明の第6の実施の形態の音声認識装置を示し、図4に示した実施の形態4の構成と異なるのは、音声合成手段6を追加したことだけであり、同一の構成要素には同一の符号を付して重複した説明は省略する。

【0039】複数の話者からの音声信号は、複数のマイクロホンから構成されるマイクロホンアレイ7を通じて入力され、ビームフォーマ8に入る。ビームフォーマ8では、目的の音声信号を抽出して音声認識手段1、話者判定手段2及び信号方向判定手段9に渡される。信号方向判定手段9では、話者の方向を推定し、その情報を話者判定手段2に渡す。話者判定手段2では、音声信号及び推定した話者方向から特定の発話者の判定が行われる。以下の流れは、実施の形態3と同じである。話者判定手段2で判定された発話者情報は、パラメータ記憶手段3と音声合成手段6に渡される。パラメータ記憶手段3では、話者情報に合わせた認識パラメータを選択して音声認識手段1に渡す。音声認識手段1では、入力された音声信号とパラメータ記憶手段3から渡された認識パラメータを利用して音声認識処理を行い、音声合成手段6に渡す。音声合成手段6では、音声認識結果と話者情

報を統合して認識結果を合成音として出力する。最後に、音声認識手段1で認識処理された結果は、パラメータ更新手段4に渡され、話者情報と認識パラメータの対応づけを更新し、パラメータ記憶手段3に渡す。

【0040】このように、本実施の形態6によれば、実施の形態4の構成に加えて、音声合成手段6を備えているので、複数の話者が発声した音声が入力された場合でも、高性能な音声認識が可能となると同時に、認識結果に話者情報を付加して合成音として出力することが可能となる。

【0041】

【発明の効果】以上のように、本発明によれば、発話者が誰であるかを分析し、その結果から各話者ごとに最適な認識パラメータを準備し、逐次最適化するような構成とすることによって、複数の話者が発声した音声が入力された場合にも高性能な音声認識を実現できるという効果を有する。

【図面の簡単な説明】

【図1】本発明の実施の形態1を示す音声認識装置のブロック図

【図2】本発明の実施の形態2を示す音声認識装置のブロック図

【図3】本発明の実施の形態3を示す音声認識装置のブ

ロック図

【図4】本発明の実施の形態4を示す音声認識装置のブロック図

【図5】本発明の実施の形態5を示す音声認識装置のブロック図

【図6】本発明の実施の形態6を示す音声認識装置のブロック図

【図7】本発明の実施の形態におけるパラメータ記憶手段に記憶されるデータ例の模式図

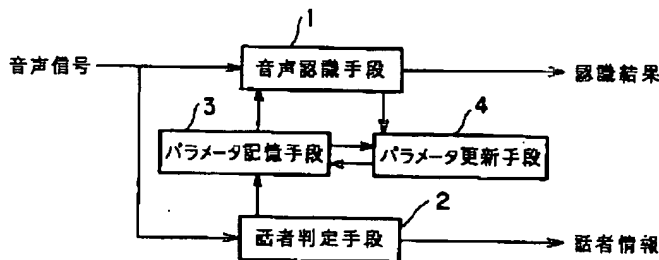
【図8】本発明の実施の形態における話者判定と認識パラメータ選択の流れ図

【図9】本発明の実施の形態における認識結果補正手段における処理例の模式図

【符号の説明】

- 1 音声認識手段
- 2 話者判定手段
- 3 パラメータ記憶手段
- 4 パラメータ更新手段
- 5 認識結果補正手段
- 6 音声合成手段
- 7 マイクロホンアレイ
- 8 ビームフォーマ
- 9 信号方向判定手段

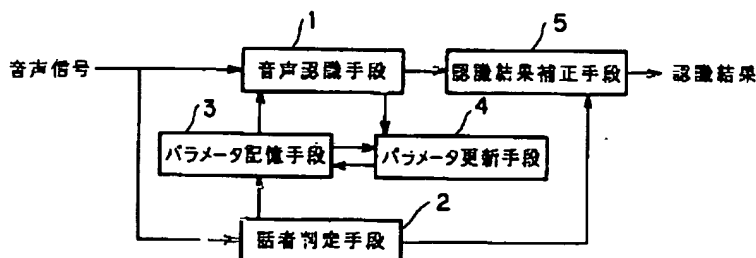
【図1】



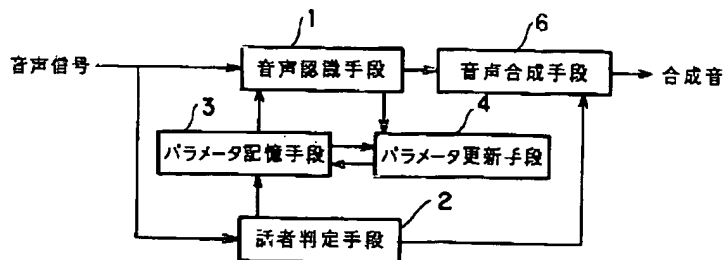
【図7】

話者情報	認識パラメータ
話者A	認識パラメータA
話者B	認識パラメータB
話者C	認識パラメータC
話者D	認識パラメータD
未登録	未登録
未登録	未登録

【図2】



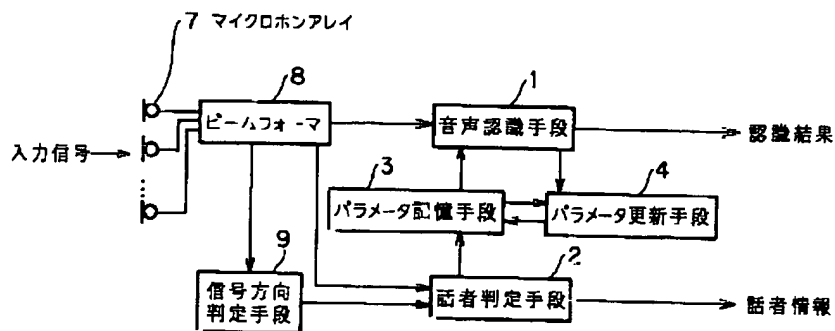
【図3】



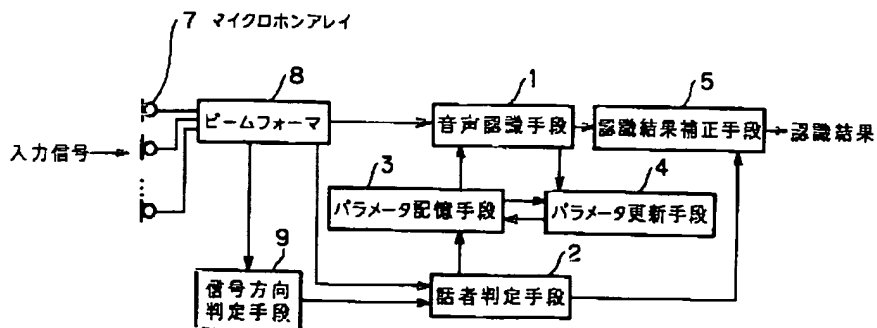
【図9】

認識の順番	話者情報の結果	音声認識結果	
		第1候補	第2候補
1	話者A	ラジオをつけて	ラジオ オン
2	話者B	次どっち?	杉戸市
3	話者A	ボウリング場	ボリウム小さく
4	話者C	今何時?	観南寺

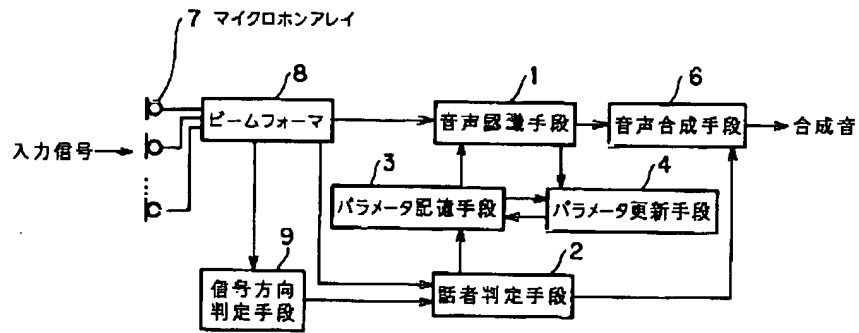
【図4】



【図5】



【図6】



【図8】

